WHAT IS CLAIMED IS:

1. A method of training a paraphrase processing system, comprising:

receiving a cluster of related texts;

selecting a set of text segments from the cluster; and

using textual alignment to identify paraphrase relationships between text in the text segments in the set.

2. The method of claim 1 wherein using textual alignment comprises:

using statistical textual alignment to align words in the text segments in the set; and

identifying the paraphrase relationships based on the aligned words.

3. The method of claim 2 wherein using textual alignment comprises:

using statistical textual alignment to align multi-word phrases in the text segments in the set; and

identifying the paraphrase relationships based on the aligned multi-word phrases.

4. The method of claim 1 wherein using textual alignment comprises:

using heuristic word alignment to align words in the text segments in the set; and

identifying the paraphrase relationships based on the aligned words.

5. The method of claim 4 wherein using textual alignment comprises:

using heuristic textual alignment to align multi-word phrases in the text segments in the set; and

identifying the paraphrase relationships based on the aligned multi-word phrases.

6. The method of claim 1 and further comprising:

calculating an alignment model based on the paraphrase relationships identified.

7. The method of claim 6 and further comprising:

receiving an input text; and

generating a paraphrase of the input text based on the alignment model.

8. The method of claim 1 and wherein selecting a set of text segments comprises:

selecting text segments for the set based on a number of shared words in the text segments.

9. The method of claim 1 and further comprising:

prior to receiving a cluster, identifying the cluster of related texts

10. The method of claim 9 wherein identifying a cluster comprises:

accessing a plurality of documents; and

identifying documents written by different authors about a common subject, as clusters of related documents.

11. The method of claim 10 wherein selecting a text segment set comprises:

grouping desired text segments of the related documents in each cluster into a set of related text segments.

12. The method of claim 11 wherein identifying documents comprises:

identifying documents written within a predetermined time of one another.

13. The method of claim 11 wherein accessing a plurality of documents comprises:

accessing a plurality of different news articles written about a common event.

14. The method of claim 13 wherein accessing a plurality different news articles comprises:

accessing a plurality of different news articles written by different news agencies.

15. The method of claim 14 wherein grouping desired text segments comprises:

grouping a first predetermined number of sentences of each news article in each cluster into the set of related text segments.

16. The method of claim 15 wherein selecting a set of text segments comprises:

pairing each sentence in a given set of related text segments with each other sentence in the given set.

17. A paraphrase processing system, comprising

a textual alignment component configured to receive a set of text segments and identify paraphrase relationships between words in the set of text segments based on alignment of the words.

18. The paraphrase processing system of claim 17 wherein the textual alignment component is configured to generate an alignment model based on statistical or heuristic alignment of the words.

19. The paraphrase processing system of claim 18 wherein the textual alignment component is configured to identify paraphrase relationships based on alignments of multi-word phrases in the set of text segments.

20. The paraphrase processing system of claim 17 and further comprising:

a clustering component configured to access a plurality of documents and cluster the documents based on a subject matter of the documents.

21. The paraphrase processing system of claim 20 wherein the clustering component is configured to cluster documents written about a same subject.

22. The paraphrase processing system of claim 20 wherein the clustering component is configured to extract predetermined text segments from clustered documents to form the set of text segments.

23. The paraphrase processing system of claim 22 and further comprising:

a pairing component configured to identify a plurality of pairs of text segments based on the set of text segments.

24. The paraphrase processing system of claim 23 wherein the pairing component is configured to identify the plurality of pairs of text segments by pairing each text segment in a given set of text segments with each other text segment in the given set of text segments.

25. The paraphrase processing system of claim 20 and further comprising:

a data store storing the plurality of documents.

26. The paraphrase processing system of claim 25 wherein the data store stores a plurality of different news articles written by different news agencies about a common event.

27. The paraphrase processing system of claim 26 wherein the clustering component is configured to cluster the news articles based on a time at which the news articles were written.

28. The paraphrase processing system of claim 27 wherein the data store is implemented in one or more data stores.

29. The paraphrase processing system of claim 17 and further comprising:

a paraphrase generator, receiving a textual input and generating a paraphrase of the textual input based on the paraphrase relationships.

30. A paraphrase processing system, comprising:

a paraphrase generator receiving a textual input and generating a paraphrase of the textual input based on a paraphrase relationship received from a textual alignment component

configured to receive a plurality of text segments and identify paraphrase relationships between words in the text segments based on alignment of the words.